Naval Research Laboratory

Washington, DC 20375-5320

NRL/FR/5740--96-9844

A Fuzzy Clustering and Superclustering Scheme for Extracting Structure from Data

JAMES F. SMITH III

Surface Electronic Warfare Systems Branch
Tactical Electronic Warfare Division

December 31, 1996

Approved for public release, distribution is unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank)*	2. REPORT DATE December 31, 1996	3. REPORT TYPE AND DATES COVERED One phase of a continuing NRL problem

4. TITLE AND SUBTITLE

A Fuzzy Clustering and Superclustering Scheme for Extracting Structure from Data

5. FUNDING NUMBERS

N0001497WX40006
AA1771319.W236

6. AUTHOR(S)

James F. Smith III

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5740--96-9844

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Arlington, VA 22217-5000

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

A fuzzy clustering algorithm based on a least-square cost function is examined. Given inorganic sensor data, the algorithm can without operator intervention decompose the data into clusters associated with particular emitters. The algorithm determines a fuzzy cluster center that represents a reduced noise estimate of measured quantities. Also, the algorithm finds the grade of membership of each data point in each cluster. The grade of membership tells the degree to which each data point belongs to a cluster, and it can be used as a measure of confidence in the cluster assignment. A second component of the algorithm superclustering allows the number of targets present in the data to be determined without a priori information, also allowing a refinement of the fuzzy cluster centers and grades of membership. The combined fuzzy clustering and superclustering algorithm is applied to both simulated and electronic support measures (ESM) data.

14. SUBJECT TERMS

Fuzzy sets
Clustering
Fuzzy logic
Inorganic sensor
Classification
Association
ESM
Electronic support measures

15. NUMBER OF PAGES

28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

CONTENTS

1. INTRODUCTION 1

2. THEORETICAL FOUNDATIONS FOR CLUSTERING, FUZZY SET THEORY, AND RELATED TOPICS 4

   2.1 Clustering, Hard and Otherwise 4
   2.2 Fuzzy Set Theory 8
   2.3 Fuzzy Clustering 8
   2.4 Picard Algorithm 9
   2.5 Defuzzification 10
   2.6 Superclustering 10

3. FUZZY CLUSTERING RESULTS FOR SIMULATED DATA 13

   3.1 Performance of the Fuzzy Clustering Algorithm 13

4. CLUSTERING RESULTS FOR DATA MEASURED BY THE ESM ATD 20

5. CONCLUSIONS 24

6. FUTURE EXTENSIONS 24

7. ACKNOWLEDGMENTS 25

8. REFERENCES 25

A FUZZY CLUSTERING AND SUPERCLUSTERING SCHEME FOR EXTRACTING STRUCTURE FROM DATA

1. INTRODUCTION

The problem considered in this report is to find structure in emitter reports from experimentally obtained inorganic sensor data. The measurement system reports data about emissions from radar units aboard both ships and aircraft. Fuzzy clustering is the approach that is considered.

Clustering is an operation that allows data to be grouped into classes defined by a similarity measure. By definition [1], given *K* objects the algorithm forms *N* clusters such that, with respect to the similarity measure, members of each cluster have a greater similarity to each other than to members of any other cluster.

This report describes a fuzzy clustering algorithm. This algorithm determines a quantity for each point and each cluster, the quantity known as the grade of membership. The grade of membership provides a measure of confidence as to how well the data are clustered. The algorithm also provides a fuzzy cluster center that represents a reduced noise value of the measured quantities. An extension of the fuzzy clustering algorithm, known as superclustering, allows the number of emitters present in the data to be determined without a priori information.

The algorithm must function in an unsupervised fashion. In other words, it requires no intervention on the part of the operator. The algorithm combines an initial operation of fuzzy clustering (FC), which creates a prespecified number of fuzzy clusters. The algorithm then uses an operation called superclustering (SC) to recombine the existing clusters, using fuzzy measures, into superclusters. The number of superclusters is optimally the number of targets. The superclusters provide data-point assignments appropriate to the targets. Superclustering has the advantage that the number of targets need not be known a priori, eliminating the need for operator intervention or interpretation. The full two-component algorithm is referred to as the fuzzy clustering and superclustering algorithm, or FCSC. Results are presented for both simulated and measured data.

The data consist of vectors, the elements of which represent measurements of various quantities: an indication of the radar type aboard the ship or aircraft being tracked (ID), pulsewidth (PW), radio-frequency (RF), pulse repetition interval (PRI), type of PRI (PRI-type), and time of intercept of the measured information. The data may include: the emitter's latitude, longitude, ellipse semi-major axis, and ellipse semi-minor axis. These data provide the input for FCSC.

Figure 1 is a histogram of real ship data with fixed ID, PW, and PRI-type as observed by an inorganic sensor system. There were 202 observations over 25 hours. Most of the observations, 90%, occurred over a 13.3-h time window. The minimum separation between non-simultaneously reported

Manuscript approved November 8, 1996.

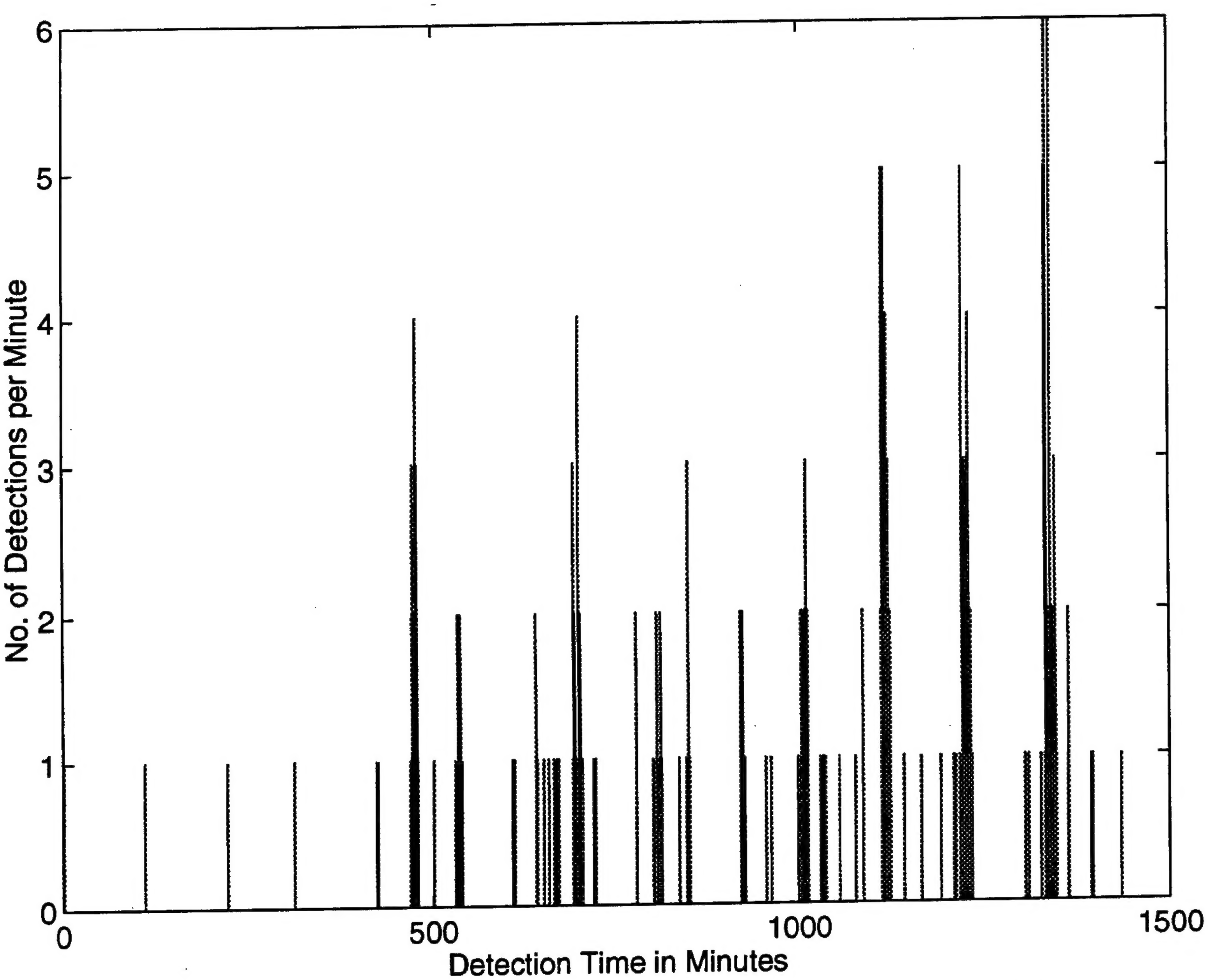

Fig. 1 — Number of ships of fixed ID, PW, and PRI-type detected by the inorganic sensor system in a 25-h time window

observations is 1 min; the maximum is 76 min, and the average is 4 min. The maximum number of observations in any minute within this window is 6. The sparsity of data found in this figure and the next is characteristic of the inorganic sensor system.

Figure 2 is a histogram of observations made by the same inorganic sensor system for an aircraft. There were 26 observations with the same ID, PW and PRI-type over a little more than 12 h. Most of the observations, about 90%, occurred over 7 h. The minimum separation between non-simultaneously reported observations is 1 min; the maximum, 111 min; and the average, 19.48 min. The maximum number of observations in any minute, within this window is 2. The highest concentration of the data are in a 4-h subwindow that contains more than 70% of the observations. Once again, data are found to be extremely sparse.

The relatively low observational density per unit time found in Figs. 1 and 2 seems to imply that data should be accumulated for seconds or minutes, at least, before beginning the analysis. The algorithm developed in this paper is discussed in terms of its batch-mode characteristics.

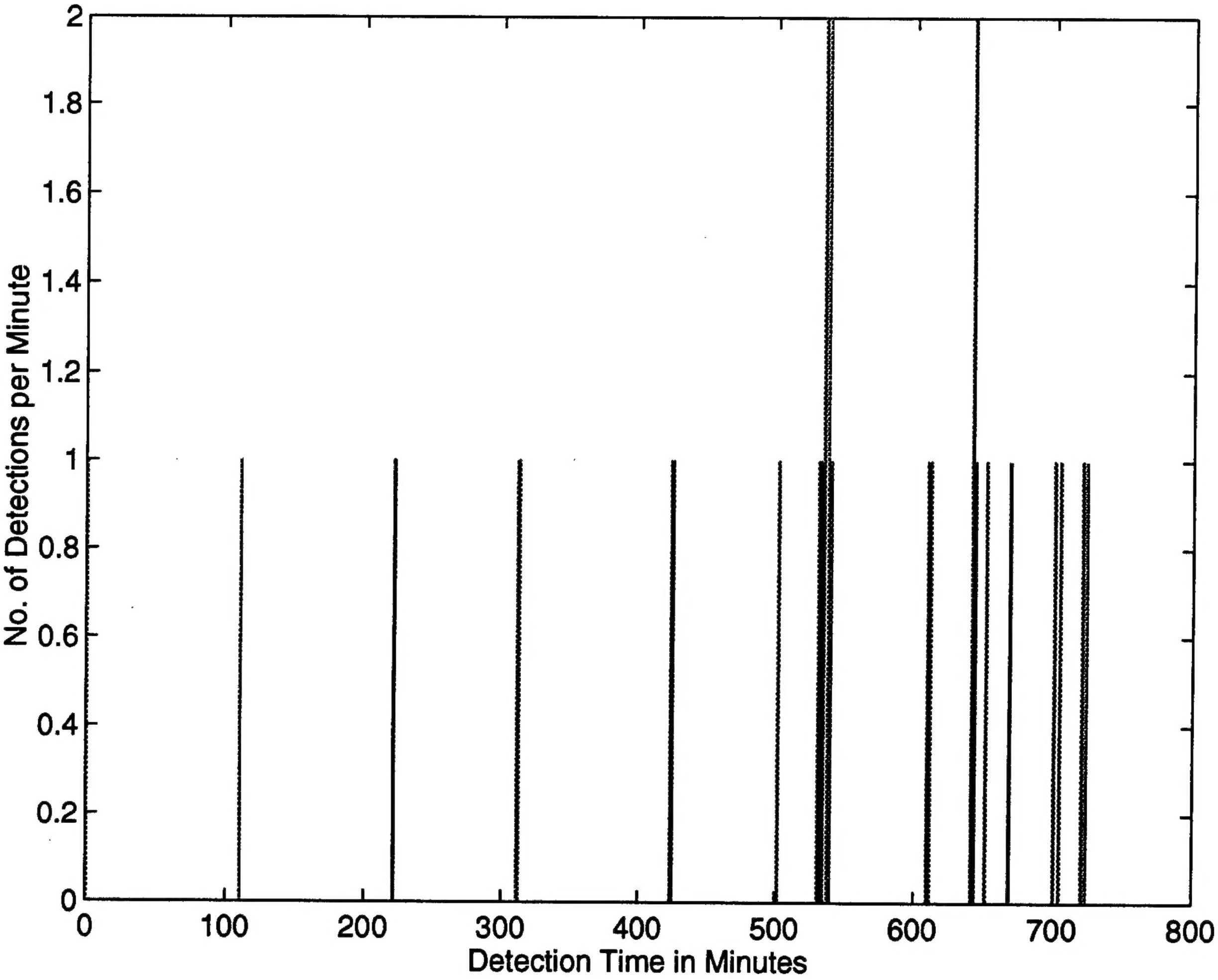

Fig. 2 — Number of aircraft of fixed ID, PW, and PRI-type detected per minute by the inorganic sensor system over about 12 h

The fuzzy clustering algorithm used here is a least-square cost function-based algorithm (Section 2.3). The cost function is minimized, resulting in the determination of the grade of membership of each data point in each fuzzy cluster and also the fuzzy cluster center. The minimization is carried out subject to the constraints that the cluster sum over grades of membership equal unity, and the related data point sum is bounded.

Clustering algorithms, including the fuzzy clustering algorithm, generally require a specification of the final number of clusters. If the data being clustered represent ships, aircraft, missiles, etc., this implies a priori knowledge of the number of targets. Obviously, the number of targets will not generally be known before processing. So it is desirable that a technique be developed to determine from the data the appropriate number of clusters, i.e., the number of targets. Such a technique, known as superclustering, has been developed that provides a solution to this problem.

Superclustering allows an improved data-point assignment and the number of emitters to be determined without operator intervention. It does this by using a fuzzy distance measure, which is the distance between fuzzy cluster centers normalized by the maximum fuzzy standard deviation for the clusters involved. There are a variety of steps, which are discussed later.

Those readers eager to see the result of applying the fuzzy clustering and superclustering algorithms to real electronic support measures (ESM) data and less interested in theory or simulation can proceed to Section 4. When the results of fuzzy clustering and superclustering of data are displayed in the bearing-time plane, the clusters appear as track-like graphs. The track-like nature of the clusters encourages comparison of the clustering results with other techniques that also produce tracks, like an Interacting Multiple Model (IMM) Kalman filter. For this reason, Section 4 also contains a plot of tracks produced by processing the same ESM data using an IMM Kalman filter.

Section 4 also discusses and illustrates a technique called product space formation subclustering. Product space formation subclustering allows results from clustering in lower dimensional parameter spaces to be used to recluster data as the parameter space dimensionality increases. This is useful for associating sensor data from multiple sensor types on different platforms. This technique allows apparent ambiguity between two emitters to be eliminated and can contribute to the recursive extension of the current batch algorithm. Product space formation subclustering should contribute to processing efficiency.

In Section 2, the concepts of hard clustering, fuzzy set theory, fuzzy clustering, defuzzification, and superclustering are introduced. Section 3 discusses fuzzy clustering results for simulated data and examines the performance of the fuzzy clustering and superclustering algorithms. Section 4 examines clustering results of data measured by ESM and national sensor systems. Section 5 provides conclusions. Finally, Section 6 discusses future extensions.

2. THEORETICAL FOUNDATIONS FOR CLUSTERING, FUZZY SET THEORY, AND RELATED TOPICS

2.1 Clustering, Hard and Otherwise

When experimental data are taken, relevant structural information should be extracted. Frequently, the underlying process that causes a spread in the data are not fully understood, so some unsupervised procedure for decomposing the data into significant subsets is desirable. Clustering represents a class of methods for solving this problem.

Clustering finds applications in many fields, and the algorithms take on many forms. Currently, powerful and popular forms of clustering use neural networks, optimization techniques, etc. This report compares two types of clustering algorithms. The first is a simple Euclidean algorithm that has proved to be useful [2]. The second algorithm is more sophisticated. It is an intuitive procedure based on fuzzy set theory that is described in more detail in subsection 2.2.

Figure 3 is a flow chart of the Euclidean clustering algorithm. The Euclidean clustering algorithm is a very simple heuristic algorithm based on the Euclidean metric. As a first step in understanding the algorithm, input and output requirements must be established. The input consists of the data to be clustered:

$\mathrm{N}_{\mathrm{min}}$ ≡ the minimum number of potential clusters;
N_{max} ≡ the maximum number of potential clusters;
r_o ≡ the initial cluster radius;

and

Δr ≡ the incremental amount the radius is increased each time the algorithm uses $N_{\max}$ clusters without fully covering the data.

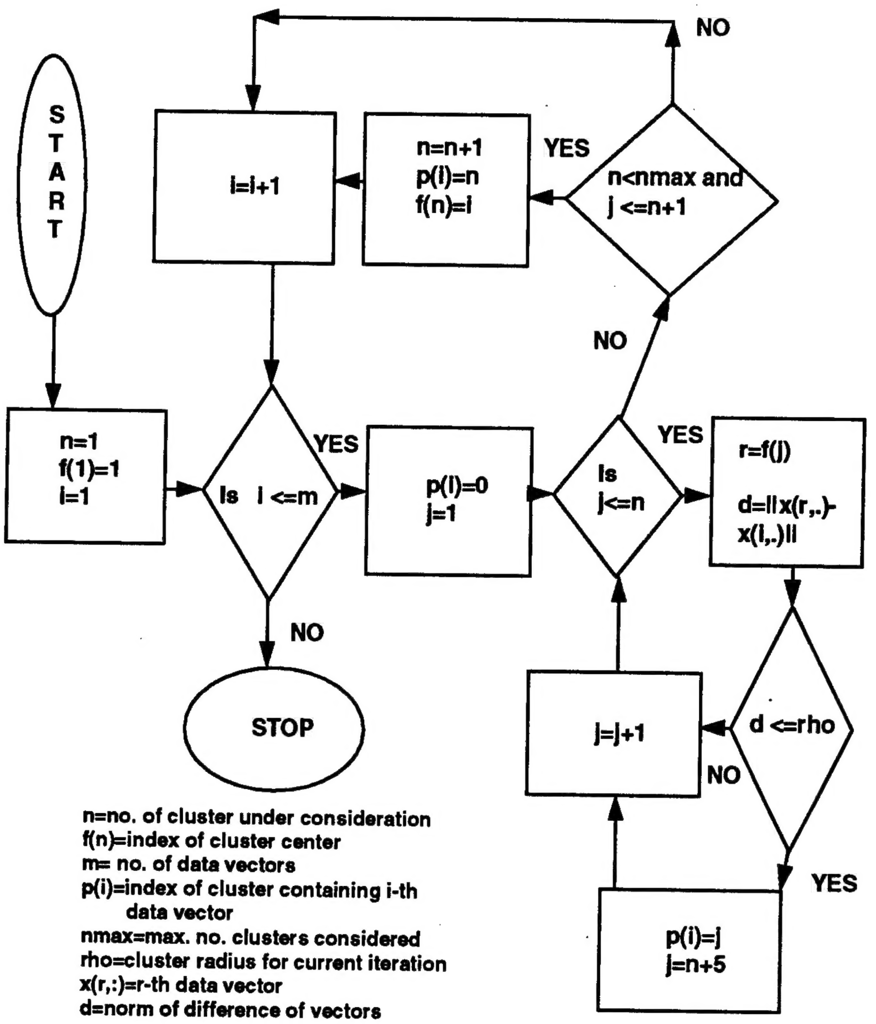

Fig. 3- Flow chart of the Euclidean clustering algorithm

When selecting the input parameters, some trial and error may be required. If, for example, clustering is to be conducted on radio frequency (RF) and pulse repetition interval (PRI) parameters, the final cluster radius is generally within a factor of three of the root-mean-square-error, $\Delta(RF,PRI)$, which is defined as:

$$\Delta(RF,PRI) \equiv \left\{\left[\frac{\Delta(RF)}{s(RF)}\right]^2 + \left[\frac{\Delta(PRI)}{s(PRI)}\right]^2\right\}^{1/2},$$

where

$\Delta(RF)$ ≡ resolution limit of RF for detector,
$\Delta(RF)$ ≡ resolution limit of PRI for detector,
$s(RF)$ ≡ RF scaling parameter,

and

$s(PRI)$ ≡ PRI scaling parameter.

The quantities $\Delta(RF)$ and $\Delta(PRI)$ are generally provided by the manufacturer of the sensor system.

When RF and PRI are independent Gaussian random variables, the root-mean-square error is the root-mean-square cluster radius. Thus, most of the data would be expected to fall within three standard deviations from the mean, as is observed in the simulated and measured cases considered in Section 3. The scaling parameters $s(RF)$ and $s(PRI)$ are used to place RF and PRI on the same scale. For the simulated and experimental cases of Section 3, $s(RF)$ and $s(PRI)$ are the standard deviations of the RF and PRI data values, respectively.

In practice, a value equal to one-tenth of $\Delta(RF, PRI)$ has been found effective for r_o and Δr. Selecting small values of r_o and Δr relative to $\Delta(RF, PRI)$ will increase the algorithm's run time. It is, however, generally a safer procedure, since the clustering algorithm is designed to select the smallest cluster radius and the smallest number of clusters that cover the data. If r_o and Δr are selected much larger than $\Delta(RF, PRI)$**,** then the number of clusters that the algorithm determines to be correct may be considerably less than the number of objects being tracked. If r_o and Δr are selected too small and $N_{\max}$ is initially sufficiently large, then each data vector will fall into its own cluster.

The selection of $N_{\min}$ and $N_{\max}$ generally also requires some trial and error. If the clustering algorithm alone were successful in extracting tracks from data, simulated or experimental, there would be one and only one track per cluster, in which case,

$$N_{\min} = N_{\max} = N_{\text{tracks}} \, ,$$

where N_{tracks} is the true number of objects being tracked.

This can be understood by again considering RF and PRI to be independent Gaussian random processes. In this case, an ideal clustering algorithm, capable of establishing cluster centers at (mean(RF), mean(PRI)) for each track and also capable of suppressing outliers, could cover most of the data with N_{tracks} clusters of radii equal to $3 \cdot \Delta(RF, PRI)$.

Rarely is clustering successful in uniquely determining the number of tracks. As such, in practice, $N_{\min}$ and $N_{\max}$ selected, so that

$$N_{\min} << N_{\text{tracks}} << N_{\max} \, .$$

The final output of this process is clustered data. The algorithm is robust under selection of different values of $N_{\min}$ and $N_{\max}$, with the number of clusters varying by at most one. Also, the assignment of data points to clusters is not changed for most points. For lower dimensional cases, one to three dimensions, results can be displayed graphically.

Having described the input and desired output, the actual operation of the program can be examined. As an initial step in the clustering operation, the first data vector in the input file becomes the center of a ball, of radius equal to the initial cluster radius. This ball is the first cluster. If the next data vector lies within this ball, it is classified as part of the same cluster. Otherwise, the algorithm forms another closed ball, centered on the second data vector. The same radius is maintained during this process. Each time the algorithm is about to form a new cluster it makes sure that the maximum number of clusters has not been exceeded. If this occurs, it creates another cluster, with the previous unclassified data vector as its center. If the maximum number of clusters has been reached, the algorithm starts over with the minimum number of clusters, first increasing the cluster radius by Δr. The algorithm continues in this fashion until all points have been classified.

In Fig. 4, the Euclidean clustering algorithm is applied to a simple example to illustrate the procedure. The left-hand column lists the maximum number of clusters that are involved in that stage of the operation. The right-hand side shows the evolution of a simulated clustering process. Clusters are shown as large circles, and there are six objects to be clustered. For this example, there is a minimum of three clusters and a maximum of four clusters. In the first step of the process, for the minimum cluster radius, the algorithm attempts to cluster by using the minimum number of clusters. The algorithm is not successful is clustering with three clusters and the given cluster radius, as can be observed, from the three points that are not covered. Having failed to cluster with three clusters, the algorithm increases the number of clusters to four, with the same cluster radius, and tries to cluster again. Again, the points are not clustered completely. Four is the preset maximum number of clusters for this example, so the algorithm again uses three clusters and increases the cluster radius by a preset amount. With this larger cluster radius, the algorithm is successful in covering the data.

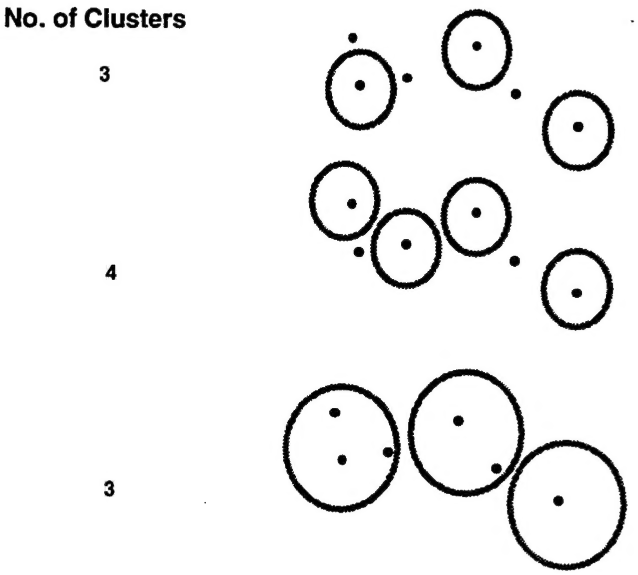

Fig. 4 — An application of the clustering algorithm to the case of six data points to be clustered, for a minimum of three clusters and a maximum of four clusters

2.2 Fuzzy Set Theory

This section provides a basic introduction to the ideas of fuzzy set theory. Fuzzy set theory allows an object to have partial membership in more than one set. It does this through the introduction of a function known as the membership function, which maps from the complete set of objects X into a set known as membership space. More formally, the definition of a fuzzy set [3] is

If **X** is a collection of objects denoted generically by $\boldsymbol{x}$,
then a fuzzy set **A** in **X** is a set of ordered pairs:

$$A = \{(x, \mu_A(x)) | \, x \in X\} \; .$$

$\mu_A(x)$ is called the membership function or grade of membership (also degree of compatibility or degree of truth) of $\boldsymbol{x}$ in $\boldsymbol{A}$, which maps X to the membership space $\boldsymbol{M}$. (When $\boldsymbol{M}$ contains only the two points 0 and 1, A is nonfuzzy and $\mu_A(x)$ is identical to the characteristic function of a nonfuzzy set.) The range of the membership function is a subset of the nonnegative real numbers whose supremum is finite. Elements with a zero degree of membership are normally not listed.

2.3 Fuzzy Clustering

The approach to clustering developed in Subsection 2.1, Euclidean Clustering, is an example of hard clustering. Fuzzy clustering differs from hard clustering in that fuzzy clustering requires a membership function to be defined, so that the grade of membership of each point within a fuzzy cluster can be established. The grades of membership will be established by minimizing a functional. This functional can be found in many places in the literature of fuzzy sets and fuzzy clustering [4,5]. It is defined below after some preliminary notation is established.

Let X be any finite set; V_{cn} is the set of real $c \times n$ matrices; c is an integer with $2 \le c \le n$, and n is the number of data points. The fuzzy c-partition space for X is the set

$$M_{fc} = \{U \in V_{cn} \, | \, u_{ik} \in [0,1] \forall i,k; \sum_{i=1}^{c} u_{ik} = 1 \forall k; 0 < \sum_{k=1}^{n} u_{ik} < n \forall i\} \; .$$

Row i of a matrix $U \in M_{fc}$ exhibits (values of) the ith membership function (or ith fuzzy subset) U_i in the fuzzy c-partition U of X. Stated less formally, u_{ij} is the grade of membership of data point j in fuzzy cluster i.

Definition: Let $J_m : M_{fc} \times R^{cp} \to R^+$,

$$Jm(U, \mathbf{v}) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m (d_{ik})^2 ,$$

where R^{cp} is the collection of possible p-dimensional vectors with real elements taken c at a time and R^+ is the real interval $[0, \infty)$;

$$U \in M_{fc}$$

is a fuzzy c-partition of *X;*

$$v = (v_1, v_2, \ldots, v_c) \in R^{cp} \qquad \text{with } v_i \in R^p$$

is the cluster center or prototype of u_i, $1 \le i \le c$;

$$(d_{ik})^2 = \| x_k - v_i \|^2 \text{ and } \| \cdot \|$$

is any inner product induced norm on R^p; and

$$\text{Weighting exponent } m \in [1, \infty) \;.$$

Since each term of J_m is proportional to $(d_{ik})^2$, J_m is a square-error clustering criteria. The solution of the fuzzy clustering problem consists of minimizing J_m as a function of U and v subject to the constraints imposed in the definition of M_{fc}. Stated more formally, solutions of

$$\min_{M_{fc} \times R^{cp}} \{J_m(U, v)\}$$

are least-square error stationary points of J_m.

The goal of the fuzzy clustering algorithm is to determine fuzzy cluster centers v_i that represent the average value of quantities in the fuzzy clusters, and the grade of membership of the *k*th data point in the *i*th fuzzy cluster for all data points-k and clusters-i. The algorithm determines these quantities by minimizing a least-square cost function where each term is weighted by a power of the grade of membership. Each term of the cost function simultaneously measures the distance of the data point from a cluster center and is weighted by the point's membership in that cluster. The minimization is conducted subject to the constraints that the sum of the grades of membership over clusters for a particular data point must equal unity, and the sum of grades of membership over data points must be bound between 1 and the number of data points for each cluster.

As input, the fuzzy clustering algorithm requires the data to be clustered, the number of anticipated clusters, and an estimate of the grades of membership or the fuzzy cluster centers. The output consists of:

- high quality estimates of the grades of membership; these quantities provide a measure of confidence of how well the data were clustered and a means of making an optimal data-point cluster assignment; and

- the fuzzy cluster centers that will represent reduced noise values of the measured quantities.

2.4 Picard Algorithm

The cost function J_m is minimized over $M_{fc} \times R^{cp}$ by using Lagrange multipliers and taking derivatives. This gives rise to a coupled iterative system of equations. When the Cauchy criterion is applied to the values of the fuzzy partition matrix, the coupled system and the Cauchy criterion are referred to as the Picard algorithm. The Picard algorithm is guaranteed to converge to a local minimum [4]. This particular type of fuzzy clustering is referred to as a c-means algorithm [5]. The relationship of the local minimum to truth depends on how much intuition the operator puts into the cost function, optimization constraints, and pre-clustering.

The system of equations resulting from the minimization represents a coupling between the fuzzy cluster centers and the fuzzy partition matrix; an initial estimate of either quantity is all that is required to initialize the iteration process. Thus, an initial estimate of the fuzzy cluster center by one class of sensors could be used to cluster data measured by another sensor system.

Another procedure for initializing the iterative process is to initially estimate the fuzzy partition matrix, i.e., the grades of membership, using some other clustering algorithm. Since the Picard algorithm is guaranteed to converge to a local minimum, if the fuzzy clustering algorithm is initialized using a good but not perfect clustering algorithm, it can frequently improve clustering results because of its ability to deal with ambiguous data-point cluster assignments.

2.5 Defuzzification

Once the data have been clustered by the fuzzy clustering algorithm, the output will be a partition matrix consisting of the grade of membership of each data point in each fuzzy cluster and the coordinates of the fuzzy cluster centers. The grade of membership gives a measure of the confidence in the data point cluster assignment and is especially useful for assigning points that fall on the boundary of two or more hard clusters. The cluster center coordinates give a convenient way of conveying the position of the fuzzy cluster. For many applications it is necessary to extract from the clustering algorithm a nonfuzzy, i.e., crisp, statement of the assignment of each point. The process of taking fuzzy results and extracting definite, i.e., crisp, data point-cluster assignments is known as *defuzzification.*

The current approach to defuzzification consists of making a definite data point assignment to the cluster for which the data point has the largest grade of membership. If a data point has equal grades of membership for more than one cluster, that point, in the simplest form of defuzzification, is assigned to the first cluster that is encountered. This is obviously arbitrary, a potentially better approach is discussed next.

2.6 Superclustering

Clustering algorithms, including the fuzzy clustering algorithm, generally require a specification of the final number of clusters. If the data being clustered represent ships, aircraft, missiles, etc., this implies a priori knowledge of the number of targets. Obviously, the number of targets will not be generally known before processing. Therefore, a technique must be developed for determining the appropriate number of clusters, i.e., the number of targets. Such a technique, known as superclustering, has been developed which provides a solution to this problem. The superclustering techniques developed here are related to and represent an extension of techniques in fuzzy cluster validity theory [5].

If the final number of clusters is not known a priori, it is usually possible to formulate a crude set of bounds on the number of clusters. For example, emissions have been detected that imply the existence of at least one emitter. On the other hand, it is known a priori that there are not more than 20 emitters. Such commonsense approaches allow bounds to be determined on the number of targets, hence clusters.

The method of superclustering is described as follows: given a commonsense bound on the number of clusters, this bound is supplied to the fuzzy clustering algorithm. The fuzzy clustering algorithm produces this number of clusters for the data, with associated grades of membership for each data point in each cluster. The fuzzy clustering algorithm also provides coordinates of fuzzy cluster centers. Intuitively, clusters should be separated—nonoverlapping and not extremely close to each other with respect to some measure. It then becomes essential to define a measure of "closeness" and provide a criterion for what "too close" means.

An obvious candidate for a measure of closeness of two clusters is the separation of the cluster centers, but the cluster centers do not tell the whole story. The data points may be distributed close to the cluster center or they may be a significant absolute distance from it. Also, when dealing with fuzzy clustering, before defuzzification, the points generally do not belong 100% to any cluster. In an effort to provide a unitless measure of closeness and incorporate the concept of vagueness inherent in fuzzy algorithms, the distance between fuzzy cluster centers should be normalized by some function of the grades of membership. Incorporating the grades of membership, i.e., superclustering before defuzzification, has the advantage of potentially better cluster assignments for points that fall on the boundary between clusters.

One such normalized measure of cluster center separation is the c-matrix defined below. Let *v(i)* and *v(j)* be the position vectors for the fuzzy cluster centers for cluster *i* and cluster *j*, respectively, and *N* the number of data points. Then the *j*th – *i*th element of the c-matrix is

$$c(i,j) = \| v(i) - v(j) \| / \max[\mathrm{std}(i), \mathrm{std}(j)] \ , \qquad (1)$$

where

$$\mathrm{std}(k) = \sum_{i=1}^{N} u(i,k)^m * [x(i) - \mathrm{mean}(k)]^2 / \sum_{i=1}^{N} u(i,k)^m \qquad (2)$$

and

$$\mathrm{mean}(k) = \left[\sum_{i=1}^{N} u(i,k)^m * x(i) \right] / \sum_{i=1}^{N} u(i,k)^m \ . \qquad (3)$$

Equations (2) and (3) define the fuzzy standard deviation and the fuzzy mean, respectively.

The c-matrix capitalizes on the intuitive idea that cluster centers should be separated by a certain number of fuzzy standard deviations. If cluster centers are closer than this, they probably correspond to the same cluster. If it is determined that two or more clusters should be merged into a single cluster, the resulting grouping will be referred to as a *supercluster.* A criterion must be established to determine when supercluster formation is warranted. A simple criterion consists of defining a threshold τ, such that if $c(i,j) < \tau$ then clusters *i* and *j* are merged into a supercluster. A method of selecting the value of τ is discussed below.

A simple criterion for selecting the value of τ would be to first consider the elements of each cluster as points randomly distributed around some mean value. If the data has a Gaussian distribution then 98% of the points will be within three standard deviations of the mean. So as a first attempt, a value of τ = 3 is selected.

The next step involved in superclustering after c-matrix formation and establishing the threshold is determining exactly how to form superclusters, i.e., when there is more than one choice based on what has been developed up to now, what is the best supercluster formation scheme. Accordingly, four different procedures for supercluster formation are discussed below. They are referred to as Fuzzy Cluster Merger Criteria 1, 2, 3, and 4 (FCMC1, FCMC2, FCMC3, FCMC4).

The first superclustering scheme, FCMC1, can be described as follows. After fuzzy clustering, the c-matrix is formed. All fuzzy cluster centers within the threshold of fuzzy cluster center one, fall into the first supercluster. Those that are within the threshold of fuzzy cluster center two fall into the second supercluster and so on.

The above procedure, obviously, allows fuzzy cluster centers to be assigned to more than one supercluster. This is not fatal because an additional fuzzy grade of membership could be assigned allowing cluster centers to partially belong to different superclusters. The procedure that has been pursued as a first effort is to uniquely assign each cluster center to a unique supercluster. This is carried out by assigning each cluster center to the first supercluster encountered.

There are obvious problems with the above procedure. Certainly, this superclustering scheme is non-unique and most likely non-optimal, but it is conceptually simple. In that sense, it is a good first start. Another related and useful tool is the formation of a fuzzy graph.

The fuzzy graph consists of plotting points on a vertical line representing each cluster center and then drawing curves to connect two cluster centers that are within threshold of each other. The curves are labeled with the value of the c-matrix elements that connect them. Strictly speaking, to be a true fuzzy graph, the c-matrix elements should be between zero and one. That is, they should be fuzzy grades of membership. This could be easily accomplished in this case by defining the matrix elements to be

$$c'(i,j) = \begin{cases} 1 & \text{if} \quad c(i,j) \geq \tau \\ c(i,j)/\tau & \text{if} \quad c(i,j) < \tau \end{cases} .$$

The advantage of the fuzzy graph is that it easily represents all possible connections between cluster centers and, as such, all possible superclustering schemes. The fuzzy graph will prove to be a useful tool in the development of the next three superclustering procedures.

The second superclustering scheme to be considered is FCMC2. For this algorithm, a cost function C_2 is defined as

$$C_2(\wp) = \sum_{K \in \wp} \sum_{i,j \in K} c'(i,j) ,$$

where P_c is the subclass of superclustering schemes that satisfy the c-matrix constraint, $\wp$ is a particular superclustering scheme in P_c, K is a particular supercluster in $\wp$, and subscripts i and j are particular cluster centers in K. The best superclustering scheme according to FCMC2 results from minimizing the above cost function over $P_{c.}$

FCMC2 is designed to select the supercluster scheme that minimizes the total sum of c-matrix elements over each supercluster. Intercluster c-matrix elements that connect different superclusters do not contribute to the sum.

The third superclustering scheme (FCMC3) consists of selecting the supercluster that maximizes the ratio

(no. of elements in supercluster with max no. elements)/(no. of superclusters in the scheme).

This method favors a large supercluster and the minimum number of superclusters. If it were not for the c-matrix constraints, all clusters would collapse into a single supercluster.

The fourth supercluster procedure (FCMC4) involves minimizing the total fuzzy ambiguity entropy over the collection of all admissible superclustering schemes, i.e., those superclustering schemes that satisfy the c-matrix constraint. The total fuzzy ambiguity entropy $\tilde{H}$ is defined as

$$\tilde{H} = \sum_{\{A|A=cluster\}} \tilde{H}(A)$$

where

$$\tilde{H}(A) = -\sum_{\omega\in\Omega}\{\mu_A(\omega)\log[\mu_A(\omega)] + [1-\mu_A(\omega))\log[1-\mu_A(\omega)]\} .$$

3. FUZZY CLUSTERING RESULTS FOR SIMULATED DATA

3.1 Performance of the Fuzzy Clustering Algorithm

Figure 5 is the starting point for a systematic analysis of the success of fuzzy clustering as a function of a parameter that characterizes the spread of the data. Four basic data points (RF_{io}, PRI_{io}), i = 1,2,3,4 are situated on the vertices of a square with side of given length referred to as the separation. The subscript notation includes a "0" to indicate they are the basic points from which others are generated.

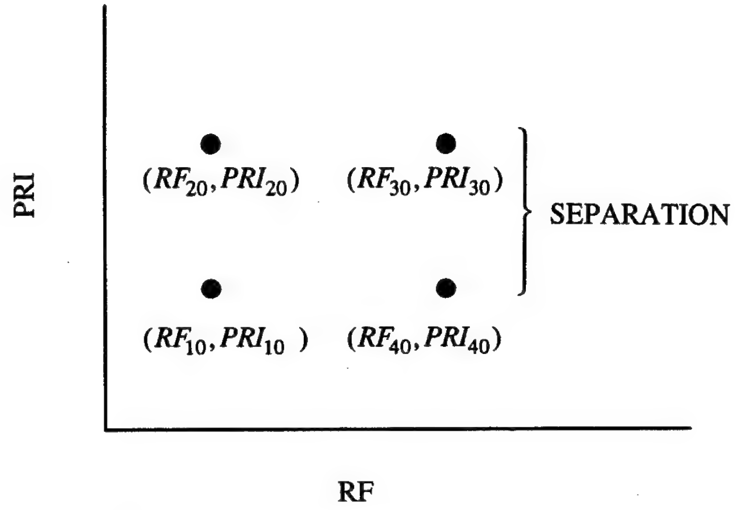

Fig. 5 — Starting point for systematic analysis of fuzzy clustering

Figure 6 presents the next step in the simulation. From the initial four points of Fig. 5 a total of 100 points are generated. The new points are generated by using a zero mean unit variance Gaussian random number generator to add points along lines parallel to the RF and PRI axis, as in Fig. 5. The random numbers for both RF and PRI are multiplied by a parameter σ. This parameter characterizes the complexity of the data. Low values of σ in the simulations that follow give rise to data groups that can be separated by drawing straight lines, i.e., *linearly separable data.* High values of σ produce data with a high degree of mixture between groups, rendering the data linearly nonseparable and more challenging for clustering algorithms.

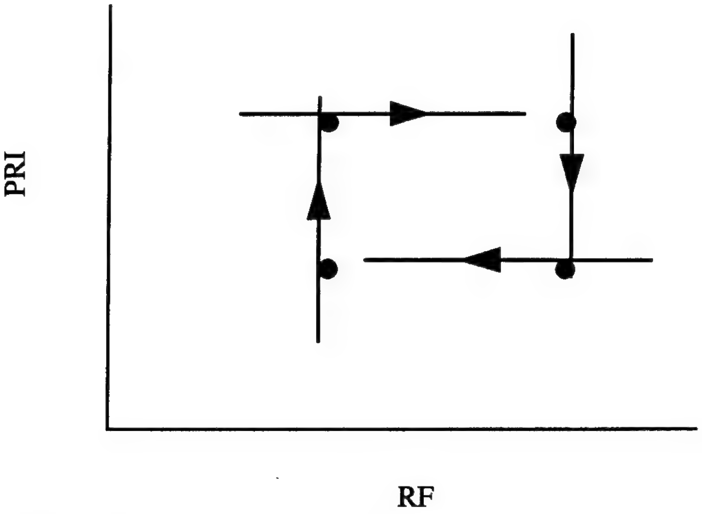

Fig. 6 — Data are simulated along fixed directions in RF and PRI plane

Figure 7 represents a systematic generation of data sets using the scheme set forth in Figs. 5 and 6. There are 12 plots, each indicating a different data set in the RF-PRI plane. Each plot is labeled by the parameter "separation/sigma." This parameter is the analog of the notion of signal-to-noise ratio. From left to right top to bottom the value of separation/sigma is increasing. Thus, the figure with separation/sigma = 0.5 represents the greatest mixing of the four clusters and hence the most difficult clustering and superclustering problem. The overlapping data in this plot preclude drawing lines between clusters, i.e., the data are not linearly separable. The figure with the least mixing of clusters is the bottom right-most figure with a value of separation/sigma = 6. The data in this plot are actually linearly separable and can be clustered by inspection. This class of data sets is designed to show that the fuzzy algorithm can still cluster when faced with a complicated noncircular data geometry with mixing of RF and PRI from various targets.

Figure 8 is a plot of the percentages of correctly clustered points as a function of separation/sigma for data generated as in Fig. 7. The upper curve represents the results for the fuzzy clustering algorithm and the lower curve the Euclidean algorithm. These are ensemble-averaging results. Each data point is the average over 64 members of the ensemble. Results are presented for a separation/sigma range of 0.5 to 3, i.e., high to low data mix limits. For this domain, the Euclidean algorithm has a success rate varying from 50% to 99% in clustering. Over the same domain, the fuzzy clustering algorithm varies from a 65% success rate at high cluster mixing to 100% of the points correctly clustered in the low data mix limit. The fuzzy clustering algorithm is always better than the Euclidean algorithm, sometimes by as much as 20%.

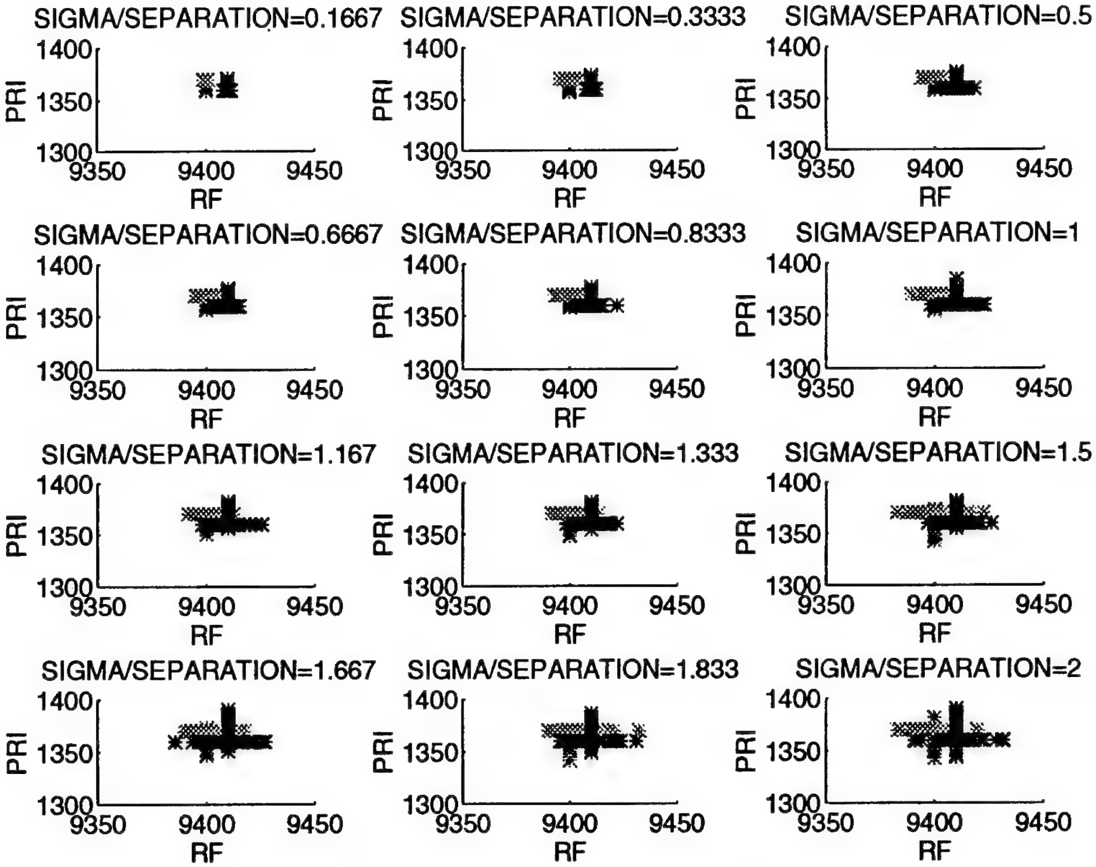

Fig. 7 — Simulated data created by adding additional points along lines parallel to RF or PRI axes

The fuzzy clustering algorithm is much more successful than the Euclidean algorithm in clustering. In particular, the fuzzy algorithm's success in clustering in the high mix limit (separation/sigma >0.5) should be noted. These are linearly nonseparable data that would be extremely difficult to cluster by visual inspection. It is also well to remember that the fuzzy clustering algorithm uses a Picard algorithm for optimization and, as such, is only guaranteed to converge to a local minimum. The current version of the fuzzy clustering algorithm uses the Euclidean algorithm to create a first estimate of the fuzzy partition matrix. If a better pre-clustering algorithm were used to initialize the fuzzy partition matrix then, unless the pre-clustering algorithm were perfect in its classification, the fuzzy clustering algorithm could most likely improve on the results.

It is also important to observe that the Picard algorithm can be initialized by selecting initial values of the fuzzy cluster centers instead of the fuzzy partition matrix. Thus, an initial estimate of the fuzzy cluster centers by one class of sensors could be used to cluster data from another class.

Another approach to fuzzy clustering would be to first pre-cluster and make an initial estimate of the cluster centers. Certain algorithms emphasize the estimate of a cluster center, notably neural net procedures like those of Linde-Buzo-Gray [6] and the Kohonen learning vector quantizer [6]. Once again, since the Picard algorithm finds a local minimum even though the algorithm of Linde-Buzo-Grey and Kohonen may produce good results on their own, the fuzzy algorithm could likely improve on their results.

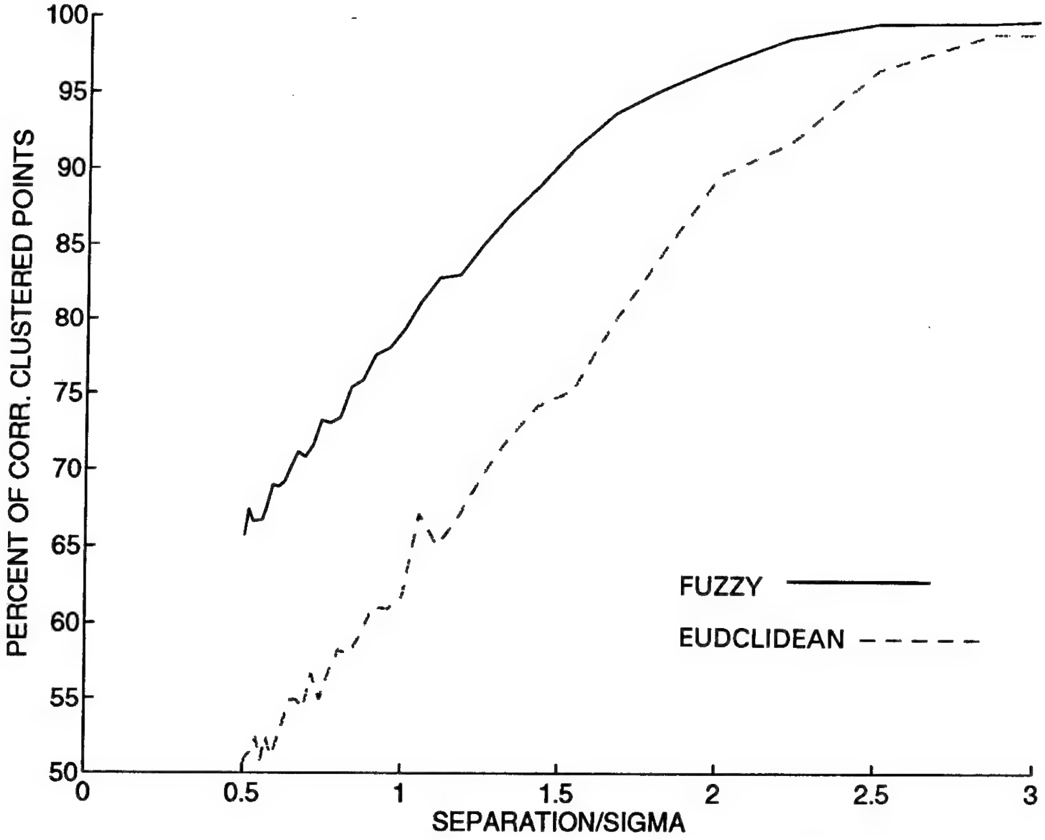

Fig. 8 — Fuzzy and Euclidean clustering results for ensemble average case

Finally, still another approach to fuzzy clustering that might improve results would be to use an optimization algorithm guaranteed to converge to a global minimum. An example of such an algorithm is simulated annealing [7]. It should be emphasized that a global minimum may not be better than the local minimum. The quality of the minima depends on the amount of a priori information contained within the cost function.

Figure 9 represents another approach to generating data for simulation. It differs from the data of Fig. 6 in terms of the data point mixing geometry.

As in Fig. 5, there are four initial points each on the vertex of a square of side of length "separation." Also, like Fig. 5, 100 points are generated by adding zero mean, unit variance Gaussian random variables multiplied by σ, which characterizes the complexity of the data. Unlike the previous case, for each data group defined by one of the initial four points, both the RF and PRI are changing. The simultaneous change of RF and PRI results in a faster mixing rate for smaller values of σ than in the previous case.

Figure 10 depicts 12 data cases generated for different values of separation/sigma, according to the procedure of Fig. 9. From upper left to lower right, separation/sigma is increasing, i.e., the data go from a high mix state to a low mix state (linearly separable data).

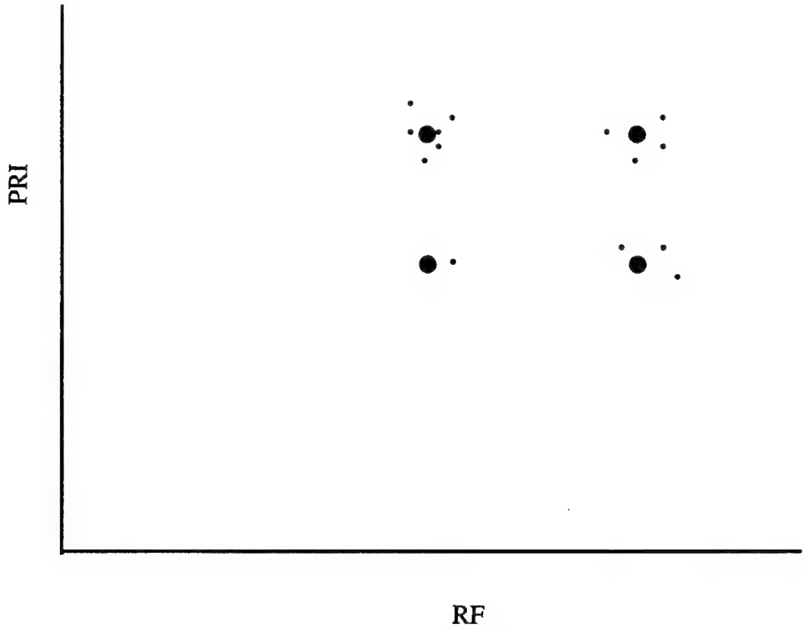

Fig. 9 — Data are simulated by generating new data points at random positions relative to the cluster centers (large dots); both RF and PRI are allowed to vary, assuming they are uncorrelated.

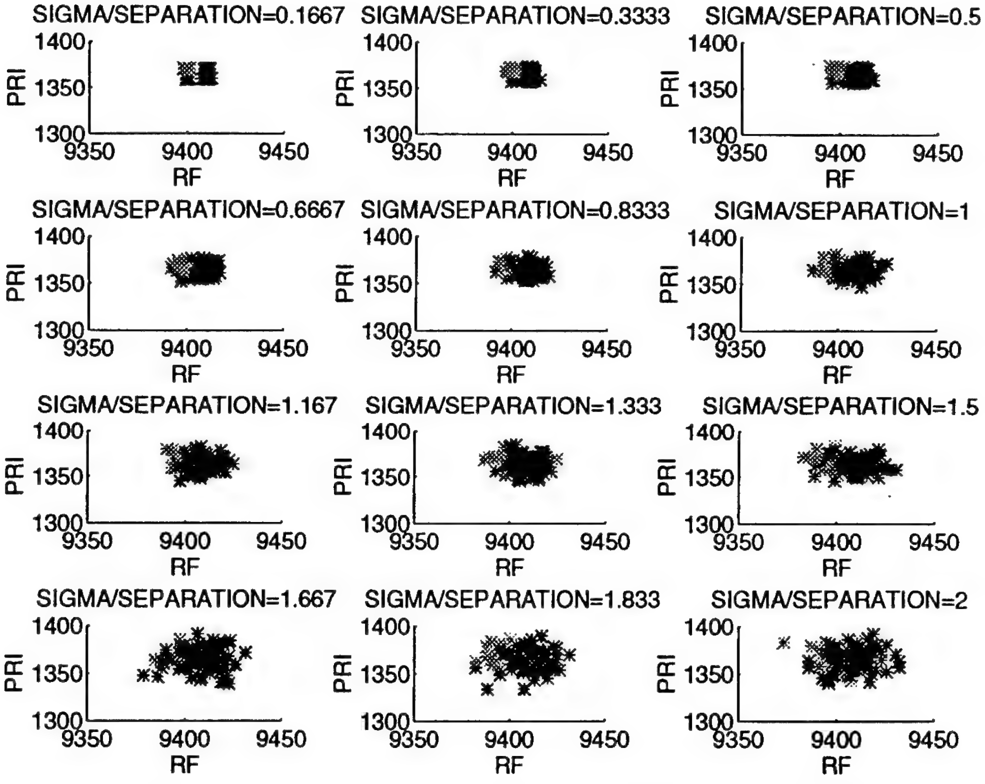

Fig. 10 — Simulated data generated by randomly changing both RF and PRI

Figure 11 is a summary plot resulting from clustering and fuzzy clustering the data cases in Fig. 10. Each data point represents an average over an ensemble with 64 members. As in Fig. 8 the separation/sigma parameter ranges from 0.5 to 3.0. The Euclidean algorithm ranges from clustering 45% of the data correctly in the high mix limit 0.5, to a 90% success rate in the low mix neighborhood. The fuzzy clustering algorithm (upper curve) is always superior to the Euclidean algorithm, by more than 25% in some cases. The fuzzy algorithm ranges from clustering 100% of the data correctly in the low mix limit to clustering half the data correctly in the high mix limit. Examination of Fig. 10 in the high mix limit suggest that without additional a priori information it is unlikely that any algorithm could automatically cluster more of the points correctly. The more complicated mixing possible in the data generation scheme of Fig. 9 results in a slightly degraded performance by both fuzzy and Euclidean algorithms.

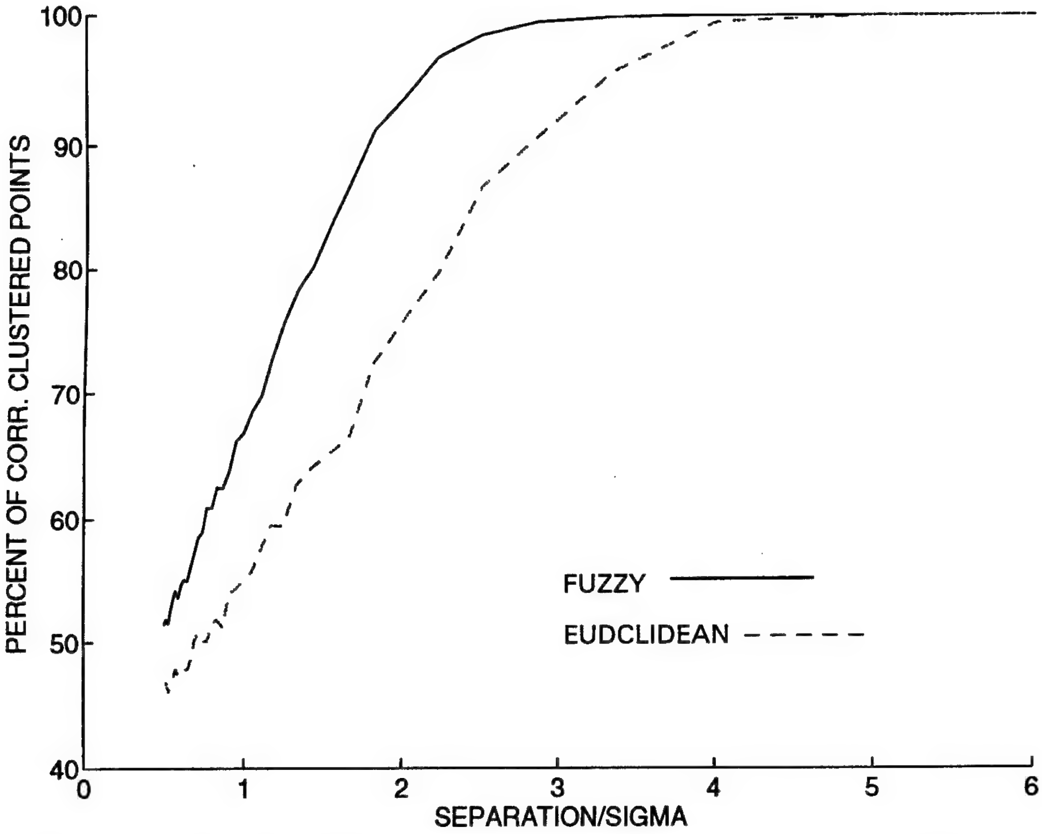

Fig. 11 — Summary plot resulting from clustering and fuzzy clustering the data cases in Fig. 10. Each data point represents an average over an ensemble with 64 members

Figure 12 examines the prediction of the cluster centers by the fuzzy clustering and superclustering algorithms and compares them to truth. The data involved four clusters of the kind given in Fig. 7. Symbols are defined as:

* – position of the true cluster centers;
o – fuzzy cluster center estimate, assuming four clusters; and
+ – supercluster estimate of the fuzzy cluster center, assuming upper bound of six clutsters.

The superclustering results are an example of no a priori information. The superclustering scheme used is FCMC4 (in this case FCMC2, FCMC3 and FCMC4 give the same results. FCMC1 gives the appropriate number of clusters but a different prediction of final cluster centers. The RF and PRI window used is 1200 to 1360 μs and 9410 to 9426 MHz.

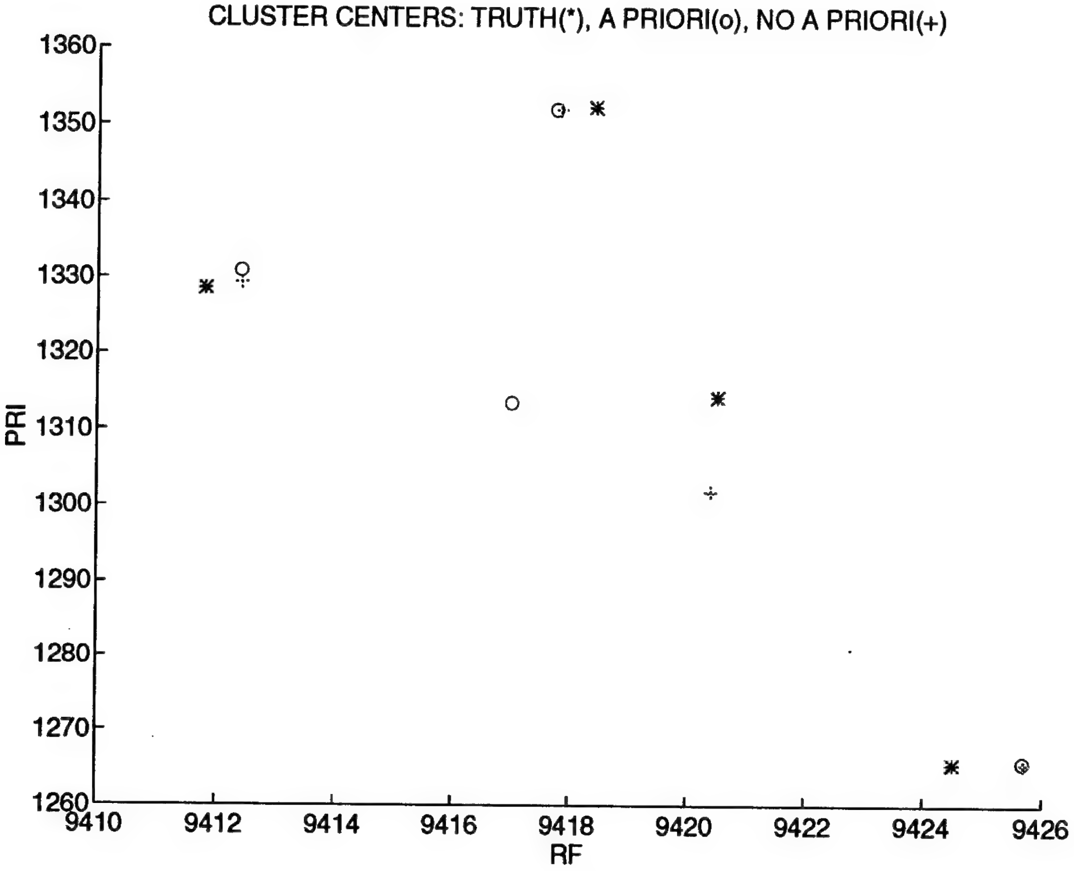

Fig. 12 — The prediction of the cluster centers by the fuzzy clustering and superclustering algorithms and compares them to truth

In each case, the fuzzy and supercluster centers are close to the true values. In three out of four cases, the fuzzy and supercluster results overlap or nearly overlap. In the fourth case, in the center of the figure the supercluster results are actually closer to truth than the initial fuzzy results. This most likely has its origin in the initializing of the Picard algorithm. The supercluster results initialized the Picard algorithm by using a combination of fuzzy cluster centers. This most likely places the initial estimate closer to a better local minimum than the original fuzzy algorithm used.

Figure 13 is the result of a superclustering ensemble calculation. The vertical axis indicates the percentage of the time the superclustering algorithm reproduced the number of emitters with an error of no more than one emitter. The horizontal axis is the standard parameter separation/sigma. This figure contains four different markers. Each marker indicates the result of a calculation for a different fuzzy cluster convergence criterion. The correspondence between symbols and fuzzy cluster merger criterions is as follows:

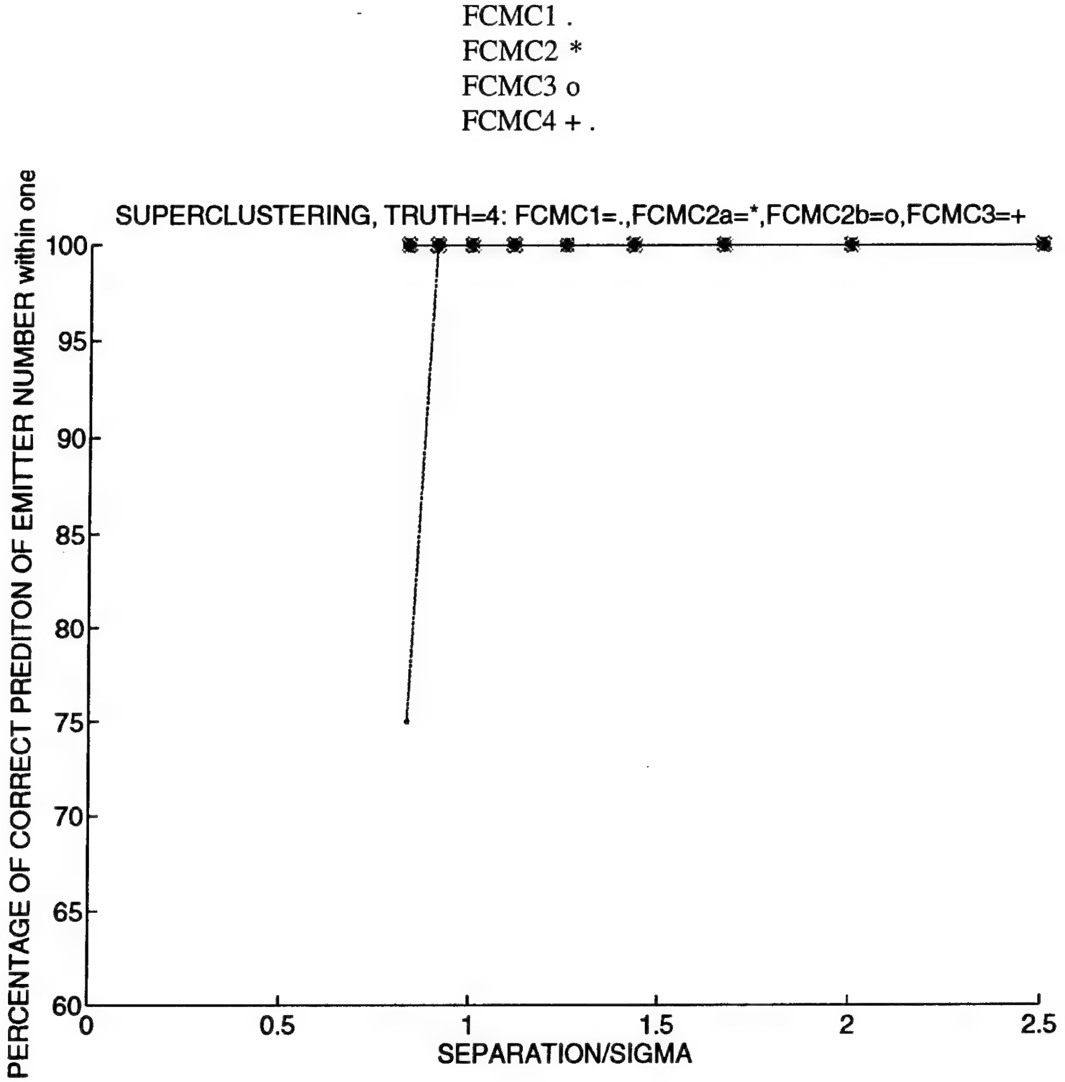

Fig. 13 — Comparison of different supercluster schemes

In this simulation, the parameter separation/sigma was allowed to vary between 0.5 and 2.5, corresponding to high to low mixing. Over the entire domain, FCMC2, FCMC3, and FCMC4 are found to determine the number of emitters within plus or minus one 100% of the time. In the high mix limit 0.5<separation/sigma<1.0, FCMC1 is off by 2 in its prediction of the number of emitters at two values of separation/sigma. Even at these points, FCMC1 predicts the number of emitters within one, 85% of the time. At the four other points in the high mix limit, it predicts the number of emitters within one 100% of the time.

4. CLUSTERING RESULTS FOR DATA MEASURED BY ESM ATD

In this section, the fuzzy clustering and superclustering algorithms are applied to data measured by a device known as the Electronic Support Measures Advanced Technology Demonstration (ESM ATD). For each emitter the ESM ATD records the following attributes: time of intercept, bearing, RF, PRI, PW, and an ID parameter. The quality of the ID parameter is uncertain, so it is generally not used in clustering.

Figure 14 represents the fuzzy clustered and superclustered data in the RF and PRI plane. The PRI is recorded on the vertical axis and it ranges from 0 μs to 1600 μs. The horizontal axis records RF that ranges from 9000 to 9900 MHz. Six clusters are present in the plane denoted by o's, *x*'s, .'s, *'s, and +'s.

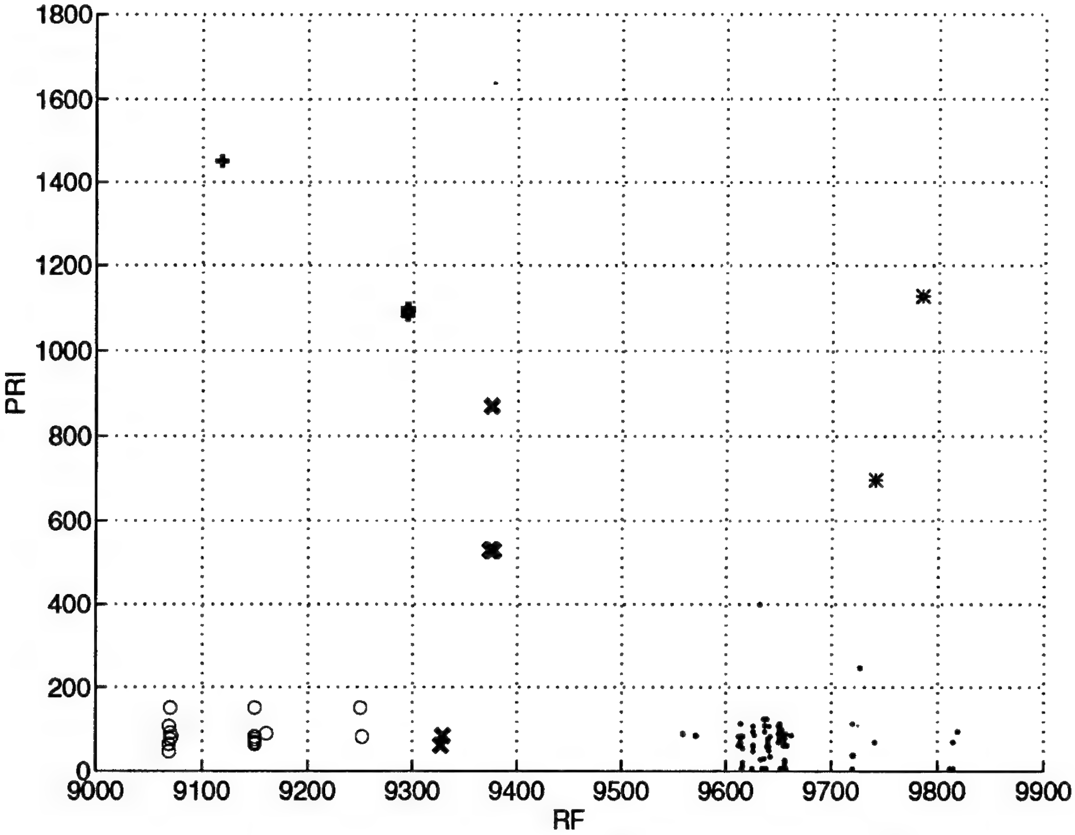

Fig. 14 — Fuzzy clustered data in the RF and PRI plane; six clusters are present

Figure 15 represents the same fuzzy clustered in the bearing-time plane. The data are confined to values of bearing between –120 to 20 deg and 66 to 78 min. The shapes of the symbols correspond to the same shape used to describe RF-PRI clusters in Fig. 14.

The process of clustering has defined clear tracks in several cases. The +'s apparently represent a target with bearing constant in time; this could be a stationary emitter. The o's indicate a target with nearly constant bearing time slope, the clustering algorithm is able to distinguish the emitter characterized by the +'s from the emitter labeled by .'s even though they cross. Several other tracks are apparent.

In Fig. 16, a PW index has been added to each data point of Fig. 15. The PW index was established by fuzzy clustering on PW within each RF-PRI cluster. The clusters denoted by points with a common symbol and index are referred to as (RF,PRI)-PW clusters. The advantage of this procedure is for cases when two emitters have essentially the same PRI and RF emissions but different pulsewidths, they can be distinguished. This is an example of a powerful technique that is referred to as *product space formation subclustering.* Clustering in lower dimensional spaces followed by forming product spaces frequently has advantages in terms of computational efficiency, easy introduction of intuitive rule sets, and a priori information.

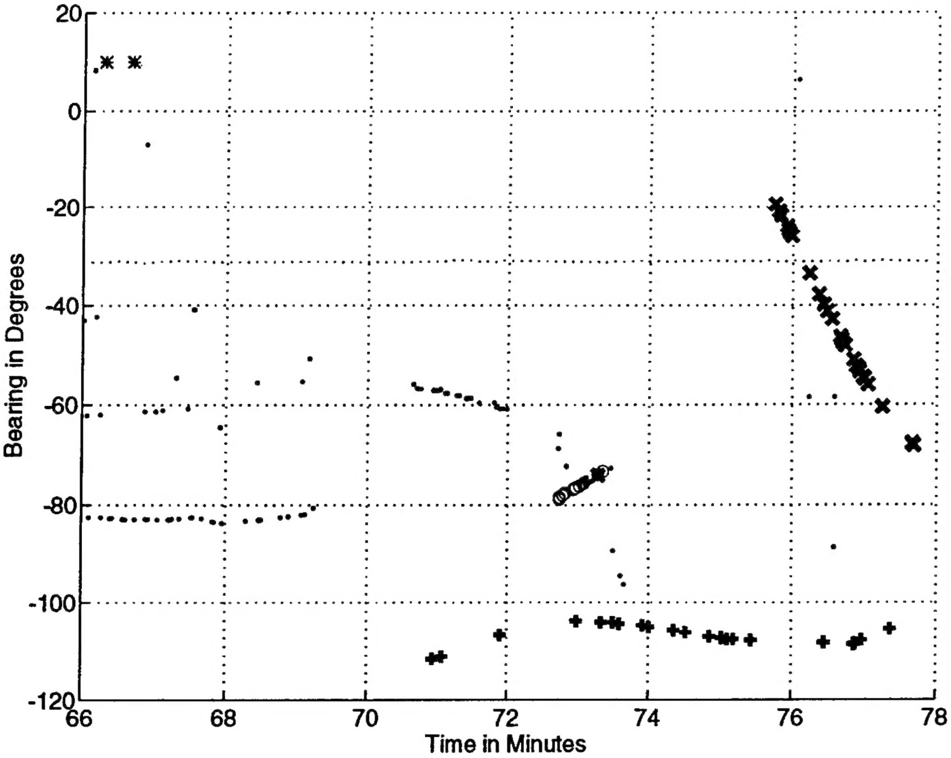

Fig. 15 — Fuzzy clustered data in the bearing-time plane; data are confined to values of bearing between –120 to 20 deg and 66 to 78 min

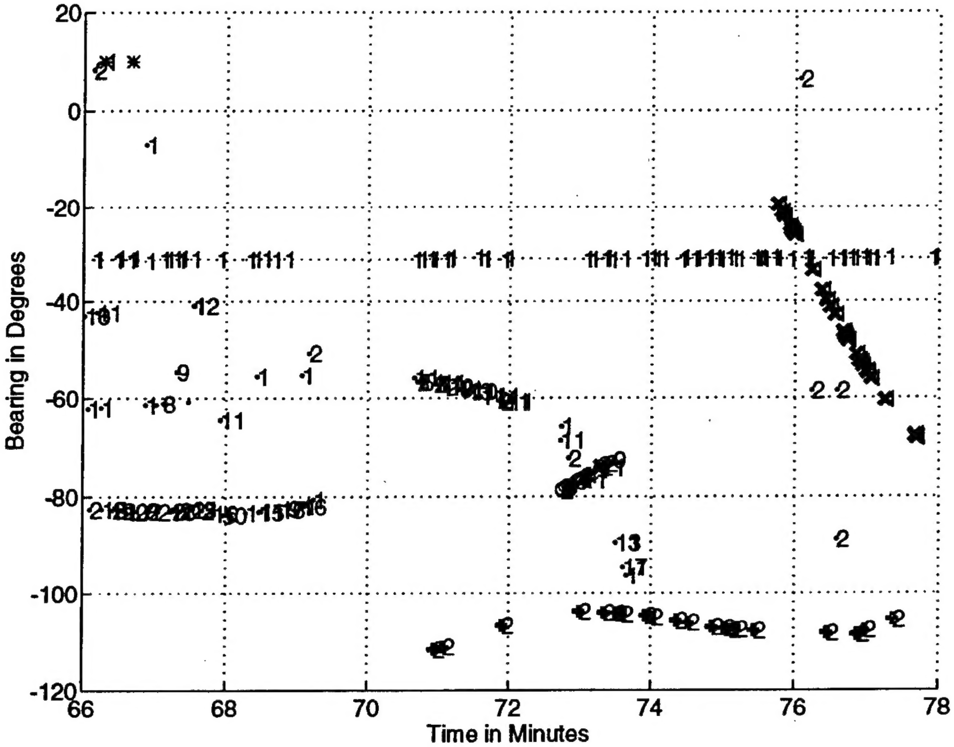

Fig. 16 — A PW index added to each data point of Fig. 15; the PW index was established by fuzzy clustering on PW within each RF-PRI cluster

The additional PW index indicates that although it might be concluded from Fig. 15 that there is a single aircraft creating tracks, in reality there are multiple aircraft. The emitter represented by +'s seems to be a single object, as are the emitters labeled by o's. The emitters labeled by o's probably represent two or more objects or one emitters rapidly changing PW.

Figure 17 is a representation of (RF,PRI)-PW clusters as tracks. Points in a (RF,PRI)-PW cluster that represent a bearing rate of more than 3 deg/s or are separated by more than 2 min without intermediate data are not connected. The algorithm has found at least six tracks. Since two or more tracks may occupy the same space, there can be additional tracks that are not immediately obvious.

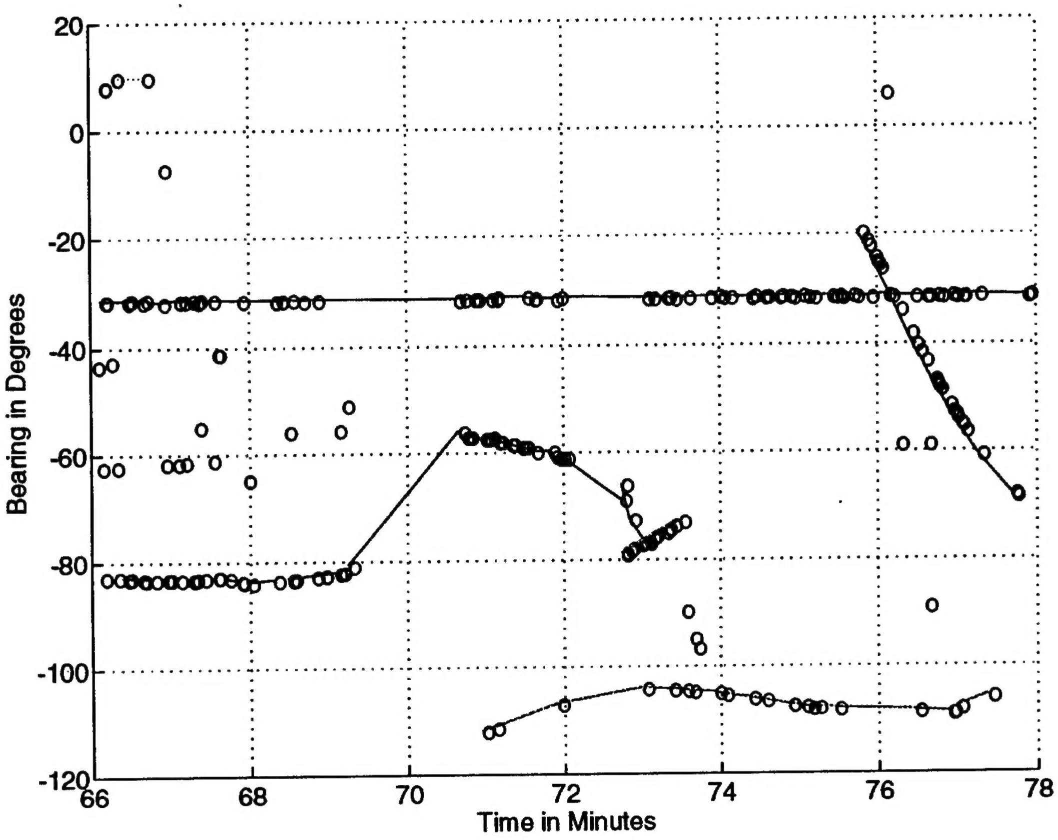

Fig. 17 — A representation of (RF,PRI)-PW clusters as tracks

Figure 18 represents the same data window processed by an IMM Kalman filter [8,9]. This filter takes into account RF, PRI, PW, bearing, and time. It also incorporates a bearing rate discrimination rule, coasting windows, and a track initialization window. Most of the same tracks that are found in this figure are also found in Fig. 17. In those instances where the Kalman filter predicts more tracks than found in Fig. 17, multiple PW indices are found by looking at the comparable cluster in Fig. 16. This indicates that the additional tracks result from including PW information.

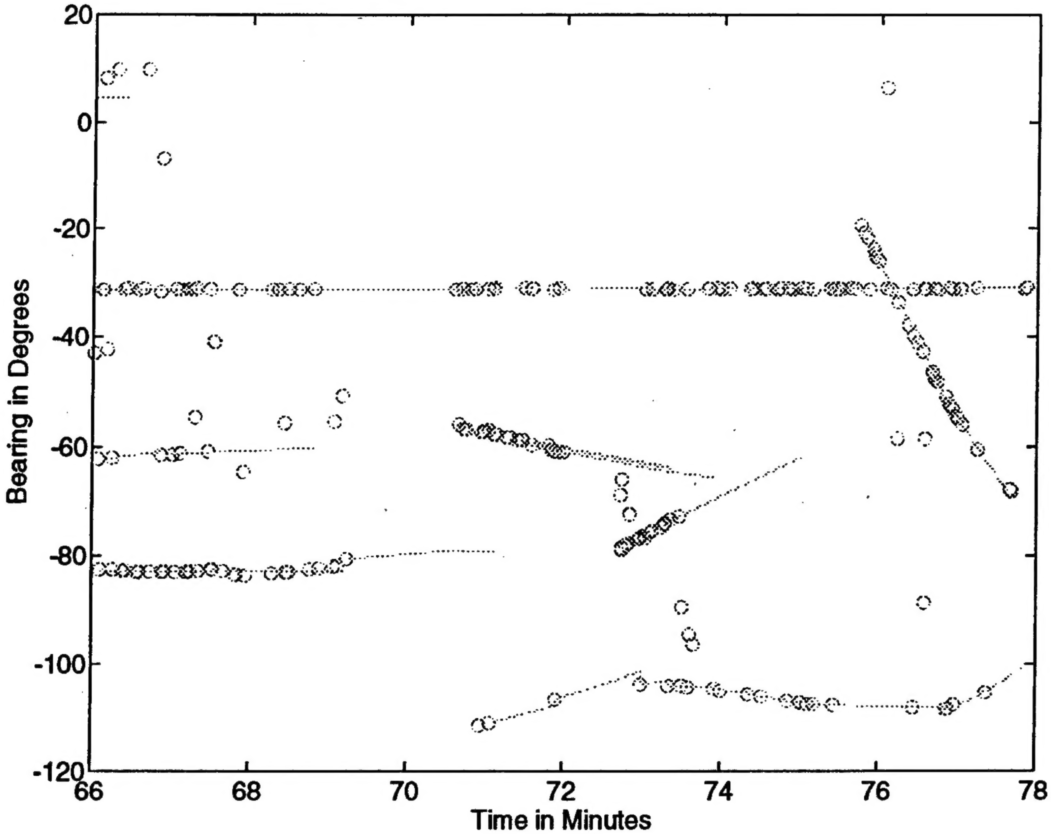

Fig. 18 — The same data as in Fig. 17 processed by a Kalman filter. This filter takes into account RF, PRI, PW, emitter ID, bearing, and time and also incorporates a bearing rate discrimination rule.

5. CONCLUSIONS

The fuzzy algorithm clusters as much as 80% to 100% of the data correctly in the low ($3 \cdot \sigma$) to high ($1 \cdot \sigma$) mix limit. Its performance exceeds that of the Euclidean algorithm by 20% or more in some cases. The fuzzy clustering algorithm with a priori knowledge of the number of clusters reproduces pre-noise addition values of attributes very well. Superclustering reproduces the number of emitters, with little or no a priori knowledge, within one. Cluster centers produced by superclustering approach the positions of those produced by fuzzy clustering with a priori knowledge of the number emitters. Superclustering determines the number of emitters without a priori information.

6. FUTURE EXTENSIONS

The current batch algorithm is being converted to a recursive real-time code. When this is accomplished, the following will be examined:

- the use of product space formation subclustering to incorporate late-arriving data from different sensor types;
- introduction of heuristic rule sets; and
- calculation of track probabilities from the grades of membership determined by fuzzy clustering.

Also, fuzzy set theoretic techniques as described above will be applied to improve an existing IMM-Kalman filter that currently uses a maximum likelihood test for association. A standard maximum likelihood test might be improved by introducing a fuzzy extension of maximum likelihood theory. Also, there may be room for application of the fuzzy extension of Bayesian theory, or Dempster-Shafer theory. Fuzzy clustering might allow the IMM-filter's current approach to multiple hypothesis track splitting and pruning to be improved or eliminated. Finally, by merging the best features of the fuzzy and Kalman algorithms, it should be possible to produce a fast algorithm that more easily incorporates subjective rule sets and uncertain information.

7. ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research. The author gratefully acknowledges useful discussions with Mr. E.N. Khoury and Dr. Joseph Lawrence III. The author also thanks Laurette Patterson, Gay Polk, Ruth Smith, and Leroy Crispell for administrative help.

8. REFERENCES

1. H. Spath, *Cluster Analysis Algorithms for Data Reduction and Classification of Objects* (John Wiley and Sons, New York, 1980).

2. J.F. Smith, A. Huynh and M. W. Kim, "An Application of Clustering and Speed Discrimination to Tracking," NRL/FR/5740--95-9801, Dec. 1995.

3. H.J. Zimmerman, *Fuzzy Set Theory and Its Applications* (Kluwer Academic Publishers Group, Boston, 1991), p. 11.

4. J.C. Bezdek and J.C. Dunn, "Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions," *IEEE Trans. Comp.* **C-24**, 835-838 (1975).

5. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Plenum Press, New York, 1981).

6. W.M. Tolles et al., "Neural Networks: An NRL Perspective," NRL/MR/1003--93-7396, Sept. 1993.

7. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in Fortran, The Art of Scientific Computing*, Second Edition (Cambridge University Press, 1992).

8. E.N. Khoury, Naval Research Laboratory, private communication.

9. Y. Bar-Shalom and X. Li, Estimation and Tracking: Principles, Techniques, and Software (Artech House, Inc., Boston, MA, 1993).